

Катедри за рачунарство и информатику
Математичког факултета у Београду
Број: и1024/2024

На седници Наставно-научног већа Математичког факултета, одржаној 3.7.2026. године, одређени смо за чланове комисије за преглед и одбрану мастер рада

Класификација типова коронавируса на основу карактеристика протеина представљених структурним алфабетом

кандидата **Николе Стојановића**. После прегледа рукописа који је доставио кандидат **Никола Стојановић**, подносимо Катедри за рачунарство и информатику следећи

ИЗВЕШТАЈ

Кандидат Никола Стојановић је доставио комисији текст свог рада. Текст се састоји из пет глава, додатка и списка референци у укупном обиму од 76 страна. У првој, уводној глави описани су коронавируси и значај њиховог разликовања и класификације. Дефинисан је и циљ рада – испитивање могућности класификације типова коронавируса на основу структурних карактеристика спајк-протеина представљених помоћу структурног алфавета Протеински блокови. У уводу је дат и кратак преглед осталих глава рада. У другој глави изложене су биолошке и структурне основе неопходне за разумевање проблема. Описани су протеини и нивои њихове структуре, структурни алфавети, протеински блокови, као и основни појмови о вирусима, коронавирусима и спајк-протеину. Посебно је описан структурни алфавет Протеински блокови који омогућава да се локална просторна организација протеина представи секвенцом симбола за 16 протеинских блокова. У трећој глави су описани коришћен скуп података, поступак припреме података, проблем небалансираности класа, коришћени алгоритми класификације (дрвета одлучивања, K најближих суседа, Наивни Бајесов класификатор, неуронске мреже, насумичне шуме, XGBoost, логистичка регресија, метод потпорних вектора) и мере за оцену перформанси класификатора. Полазни скуп чине експериментално одређене структуре спајк-протеина које су преузете из базе RCSB Protein Data Bank. Скуп је допуњен структурама мање заступљених типова коронавируса које су предвиђене помоћу алата AlphaFold2/ColabFold. Алат PVPlore је коришћен за одређивање секвенци протеинских блокова за издвојене спајк-протеине. Из секвенци су издвојене учесталости појединачних протеинских блокова, биграма, триграма и њихове комбинације за потребе класификације. Експериментални скуп садржи 1027, а проширени 1325 инстанци осам типова коронавируса. У четвртој глави приказани су и анализирани резултати истраживања на експерименталном и проширеном скупу података, за различите скупове карактеристика. Најбољи резултат на експерименталном скупу постигнут је логистичком регресијом над комбинацијом учесталости протеинских блокова и биграма, са макро F1-мером 0,86, док је на проширеном скупу логистичка регресија над триграмима постигла макро F1-меру 0,94. Резултати показују да n-грамске карактеристике дају боље резултате од самих учесталости протеинских блокова и да проширивање скупа предвиђеним структурама доприноси бољој класификацији мање заступљених типова коронавируса. У петој глави дат је закључак рада, док су у додатку приказани извештаји о класификацији по класама за све коришћене скупове карактеристика и све разматране алгоритме.

ЗАКЉУЧАК

Предлажемо Катедри за рачунарство и информатику да рад ***Класификација типова коронавируса на основу карактеристика протеина представљених структурним алфабетом*** кандидата **Николе Стојановића** прихвати као мастер рад.

У Београду, 14.9.2026. године.

КОМИСИЈА:

Мирјана Маљковић Ружичић, ментор

Јована Ковачевић

Саша Малков
